Proceduralizing control and discretion : Human oversight in artificial intelligence policy

Koulu, Riikka

2020-12-01

Koulu , R 2020 , ' Proceduralizing control and discretion : Human oversight in artificial intelligence policy ' , Maastricht Journal of European and Comparative Law , vol. 27 , no. 6 , pp. 720-735 . https://doi.org/10.1177/1023263X20978649

http://hdl.handle.net/10138/326951
https://doi.org/10.1177/1023263X20978649

Article

MJ

Maastricht Journal of European and Comparative Law
2020, Vol. 27(6) 720–735

DOI: 10.1177/1023263X20978649
maastrichtjournal.sagepub.com

Proceduralizing control and discretion: Human oversight in artificial intelligence policy

Riikka Koulu*

Abstract

This article is an examination of human oversight in EU policy for controlling algorithmic systems in automated legal decision making. Despite the shortcomings of human control over complex technical systems, human oversight is advocated as a solution against the risks of increasing reliance on algorithmic tools. For law, human oversight provides an attractive, easily implementable and observable procedural safeguard. However, without awareness of its inherent limitations, human oversight is in danger of becoming a value in itself, an empty procedural shell used as a stand-in justification for algorithmization but failing to provide protection for fundamental rights. By complementing socio-legal analysis with Science and Technology Studies, critical algorithm studies, organization studies and human-computer interaction research, the author explores the importance of keeping the human in the loop and asks what the human element at the core of legal decision making is. Through algorithmization it is made visible how law conceptualises decision making through human actors, personalises legal decision making through the decision-maker's discretionary power that provides proportionality and common sense, prevents gross miscarriages of justice and establishes the human encounter deemed essential for the feeling of being heard. The analysis demonstrates the necessary human element embedded in legal decision making, against which the meaningfulness of human oversight needs to be examined.

Keywords

Algorithmic decision making, AI ethics, AI regulation, automation, human oversight, EU policy

Introduction

Following the recent advancements in artificial intelligence (AI), algorithmic decision making (ADM) systems are being increasingly deployed to support or completely automate legal decision

* University of Helsinki, Helsinki, Finland

Corresponding author:
Riikka Koulu, University of Helsinki Legal Tech Lab, Faculties of Social Sciences and Law, University of Helsinki.
E-mail: riikka.koulu@helsinki.fi

making across the public domain, including courts and public administration. Defined as encoded procedures for solving problems by transforming input data into a desired output and producing recommendations on this basis,[1] algorithmic systems are said to contribute to the 'algorithmisation' of governance, a distinct form of social ordering that becomes entwined with autonomous software.[2] A range of ADM systems are used across our societies to facilitate or automate decision making, examples ranging from online activities such as curation of search engine results, targeted advertising and content moderation to organisational processes such as recruitment decisions, managerial surveillance and resource allocation.

Algorithmic systems also increasingly contribute to decisions on public administration, whether a person is entitled to a social benefit, whether a family will be in need of child protection services, or whether an immigrant gets refugee status or citizenship, with great hopes for AI deployment in the judiciary also being expressed.

The concern for fundamental rights has created a global push towards AI ethics, producing a plethora of ethical guidelines meant to limit the risks and negative consequences associated with the algorithmisation of society.[3] Sometimes framed as 'ethics washing', the instruments have been criticised for their non-binding nature, blurry scope of application, and lack of clear implementation guidelines for programmers and administrators of justice, that could be implemented easily through checklists, all of which contributes to their limited ability to regulate AI systems.[4] However, the problem representations as well as the solutions proposed are bound to influence the emergent hard-law approaches, as the juridification of AI regulation proceeds.[5]

Currently, human oversight is advocated by a range of actors as a focal ethical principle for AI development and deployment. For example, the EU Commission's Communication in 2019 portrayed human agency and oversight as the first of seven key requirements AI applications must follow to be considered trustworthy.[6] The risks and challenges hoped to be addressed by human oversight include dangers to human autonomy, lack of transparency and opaque algorithmic models, privacy and data protection issues, as well as discrimination.[7] Similarly, the ethical guidelines developed by the Council of Europe's European Commission for the efficiency of justice

1. T. Gillespie, 'The Relevance of Algorithms', in T. Gillespie, P. J. Boczkowski and K. A. Foot (eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (MIT Press, 2014), p. 167.
2. A. Aneesh, 'Global Labor: Algocratic Modes of Organization', 27 *Sociological Theory* (2019), p. 347; K. Yeung and M. Lodge (eds.), *Algorithmic Regulation* (Oxford University Press, 2019).
3. The German non-profit organisation AlgorithmWatch provides online a global inventory of AI ethics Guidelines listing over 80 documents at the time of writing in April 2020. See, 'AI Ethics Guidelines Global Inventory', *AlgorithmWatch* (2020), https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/.
4. See e.g. T. Hagendorff, 'The Ethics of AI Ethics: An Evaluation of Guidelines', 30 *Minds and Machines* (2019), https://arxiv.org/abs/1903.03425; B. Mittelstadt et al., 'The Ethics of Algorithms: Mapping the Debate', 3 *Big Data & Society* (2016), p. 1; D. Greene, A. L. Hoffmann, and L. Stark, 'Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning', *Proceedings of the 52nd Hawaii International Conference on System Sciences* (2019), http://hdl.handle.net/10125/59651.
5. It should be noted that for the time being the regulatory landscape regarding the use of ADM systems in the society at large or in the legal domain remains unclear, although the urgent need for socio-legal research is widely acknowledged. See K. Yeung and M. Lodge (eds.), *Algorithmic Regulation* (Oxford University Press, 2019); J. Cohen, *Between Truth and Power* (Oxford University Press, 2019).
6. Communication from the Commission to the European Parliament, the Council, the European economic and social committee and the Committee of the regions building trust in human-centric artificial intelligence, COM/2019/168, p. 3.
7. See e.g. R. Koulu, 'Human control over automation: AI ethics and EU policy', 12 *European Journal of Legal Studies* (2020), p. 9.

(CEPEJ) emphasise the need for user control over AI and that AI deployment should not undermine access to a judge, going as far as suggesting a right to a natural judge.[8]

In this article, I examine human oversight from a socio-legal perspective, focusing particularly on the procedural dimension of algorithmisation in the context of legal decision making. Legal decision making in public administration and in the judiciary constitutes a specific context for AI deployment, characterised by a high level of regulation that aims to provide procedural safeguards as well as accountability mechanisms. This said, legal processes produce a variety of decisions with varying legal effects, ranging from the mundane routine-like cases in public administration to cases with wide discretionary power such as civil and commercial adjudication. Understanding this diversity brings the level of discretion at the core of examination, as the broader discretion of the human decision-maker is often perceived to also constitute an increased risk of arbitrariness.[9] Simultaneously, discretion is vital for introducing reasonability and context-sensitivity to the application of law as well as for producing systemic renewal through precedents. Hence the legal system aims to control the use of public power by striking a balance between sufficient freedom in the form of discretion and sufficient due process and accountability structures that temper that decision making power.

However, it is not only reined-in discretionary power that defines legal decision making but also the physical confrontation between the parties and the judge.[10] This encounter is at the core of our understanding of fair trial, which often assumes such an encounter in the physical manifestation of the day in court. Although not necessarily explicitly said, the encounter gives substance to many due process standards such as the right to be heard and equality of arms, and also encompasses much of the critique of the access to justice and ADR movements that emphasised the practical and emotional needs of those seeking justice instead of formal institutional settings and perspectives.

Algorithmisation of legal decision making raises questions about both of these concepts and the answers may reveal something fundamental about the ways law defines fairness. What happens to the decision-maker's discretion with increasing reliance on ADM systems? Can you encounter a machine in a meaningful way? These questions define what ADM deployment means for legal decision making, the decision-maker's moral and legal responsibility over machine-generated decisions and meeting the due process and justice expectations of those affected. In other words, law assigns the power and the responsibility to the human decision-maker for the realization of due process, substantial justification and institutional legitimacy of legal decision making.

8. See, European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, *CEPEJ* (2019), p. 8, 15, https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment. Interestingly, the annexed study connects the right to a judge to natural persons: 'There is also a need to consider whether these solutions are compatible with the individual rights enshrined in the European Convention on Human Rights (ECHR). These include the right to a fair trial (particularly the right to a natural judge established by law, the right to an independent and impartial tribunal and equality of arms in judicial proceedings) and, where insufficient care has been taken to protect data communicated in open data, the right to respect for private and family life.' [emphasis added].
9. E.g. N. Luhmann, *Organization und Entscheidung* (VS Verlag für Sozialwissenschaften, 2000), p. 136, also p. 51 and 294.
10. For example, Vilhelm Aubert recognises in his legal anthropological research a specific dynamic of conflict management across cultural contexts, where the introduction of a neutral third as the representative of the society transforms the dyad of conflicting parties into a triad. See V. Aubert, 'Competition and Dissensus: Two Types of Conflict and Conflict Resolution', 7 *Journal of Conflict Resolution* (1963), p. 26.

The aim with this article is to draw an overview of the necessary human element we associate with legitimate and fair legal decision making. I argue that algorithmisation challenges existing conceptualisations of legal decision making by making visible the inherently human face of procedural justice, articulated both through the procedural norms regulating decision-makers' use of discretion and the procedural experiences of those seeking justice. Ultimately, law still perceives the human decision-maker as constitutive for fair decision making. Simply put, humans are the implicit medium of law. This human medium has been engrained in the deep structure of law, to the form of legal institutions and processes. Where should the line between human and machine action be drawn? What makes human input meaningful?

Although no clear line can be drawn because of the complexity and contextuality of algorithmisation, a more nuanced understanding of the human element enables us to ask what is needed from human oversight for due process. Algorithmisation crystallises the importance of humans in two aspects of legal decision making. First, the human element is embedded in the decision-makers' use of discretion. Second, the human-faced ideal is inherent in the procedural values behind the feeling of being heard and encountered as a human being for those seeking justice. Of these two aspects, discretion and encounter, this article focuses particularly on the first due to the prominence of the judge's perspective in procedural law. By bridging debates on AI ethics and algorithmisation, early research on the impact of technical systems on organisational decision making and socio-legal research, this procedural context also complements and guides critical policy analysis on algorithmic decision making and the role of the human overseer.[11] This enables us to ask what is meant by human oversight and, more importantly, what do we want it to mean.

The argument is built in four steps. First, the analysis is contextualised in terms of algorithmisation of legal decision making that calls for a systemic reassessment of current legal thought (section 2). Second, a brief description is provided of the current EU regulation and emerging AI policy, depicting the importance attributed to human oversight as a procedural safeguard against the risks of algorithmic systems (section 3). Third, based on insights from socio-legal and human-computer interaction research, the feasibility of human oversight is contested (section 4). Finally, the human element of legal decision making is located in the importance given to human decision-maker's discretionary power, the use of which algorithmisation is changing (section 5).

Algorithmisation as law's mirror: systemic risks and shortcomings of law

In this section, I briefly discuss how algorithmisation is seen as a fundamental challenge to how law operates, how it conceptualises rights and interests, and establishes actors, institutions, and procedures. By aggravating the need for systemic reassessment, algorithmisation provides a mirror for critical reflection, enabling us to elaborate the shortcomings of existing legal structures, revealing the limitations of current legal doctrine. In the context of legal decision making, such analysis touches particularly on procedural and administrative law doctrines.

Despite siloed approaches, there is a widespread consensus regarding the need to reform dispute resolution mechanisms, shared both by access to justice scholars and socio-legal algorithm

11. R. Koulu, *European Journal of Legal Studies* (2020) (Forthcoming), p. 9.

studies.[12] The shortcomings of existing court and ADR mechanisms are acknowledged and digital technologies are often suggested as a solution to many of the problems faced by publicly funded conflict management: the pressure for more efficiency and continuous budget cuts, delays in processing time and demands for improved quality, insufficient guidance for citizens, the well-being of government officials at work, and the changing expectations of society. Hailing from the US, where the civil justice crisis runs deeper than in many European justice systems, Julie Cohen points out that there is no reason why dispute resolution mechanisms should remain the same any more than there is a reason to assume their algorithmisation would lead to increased fairness or efficiency. Instead, understanding the dynamic complexity forms the starting point for reassessment:

"The ongoing processes of judicial retrenchment and reconfiguration are the products of a complex encounter between the liberal-activist paradigm underlying the traditional, court-centred system of procedural justice, the affordances that networked digital technologies offer for large-scale information aggregation and processing, and the ascendant ideology of neoliberal governmentality."[13]

If human oversight is the solution, what is the problem? Karen Yeung conceptualises problems caused by algorithmisation by differentiating between (i) process-based concerns, (ii) outcome-focused concerns and (iii) predictive personalisation of services.[14] Process-based concerns include formal due process and actual capabilities needed for contestation but also issues of transparency, explainability and reason giving, as well as the dehumanising impact automation has on decision making processes.[15] Outcome-based concerns are more substantive and may relate to discriminatory decisions and the aggravating impact ADM systems have on societal inequalities, whereas mass surveillance and behaviour modification are risks related to the predictive personalization of services.[16] Furthermore, Yeung emphasises the cumulative effect of algorithmisation that should be assessed not only from the perspectives of individuals and groups, but also from those of the overall society. Legal rights have an intergenerational scope that also obliges current generations to ensure legal protection and renewal of law for future generations, although individualistic rights-discourses often dismiss this overarching obligation. Following similar argumentation as Cohen, Yeung perceives the risks of algorithmisation as systemic and collective, which further limits the feasibility of individualistic legal doctrine to conceptualise these challenges.[17]

In a sense, algorithmisation can be seen as a step closer to the ideal-type model of Weberian bureaucracy, defined by hierarchy of power, supremacy of written rules and discipline and control, promising legal certainty to the letter.[18] This way, algorithmisation comes with the promise of improved quality by removing human shortcomings, errors and biases. At the same time, growing awareness of ADM risks and challenges demonstrates the (often implicit) importance granted to

12. See e.g. J. Resnik, 'A2J/A2 K: Access to Justice, Access to Knowledge, and Economic Inequalities in Open Courts and Arbitrations', 96 *North Carolina Law Review* (2018), p. 102; J. Cohen, *Between Truth and Power*, p. 143-144.
13. J. Cohen, *Between Truth and Power*, p. 143-144.
14. K. Yeung, 'Why Worry about Decision making by Machine?', in K. Yeung and M. Lodge (eds.), *Algorithmic Regulation* (Oxford University Press, 2019), p. 24.
15. Ibid. p. 24-31.
16. Ibid. p. 31-35.
17. Ibid. p. 42.
18. I. Koivisto, 'Thinking Inside the Box. The Promise and Boundaries of Transparency in Automated Decision making', *EUI Working Papers* AEL 2020/01 (forthcoming), https://cadmus.eui.eu/handle/1814/67272.

human decision-makers' *in casu* consideration and common sense, the application of discretionary power in a given context when necessary. As Yeung puts it, the removal of humans also removes human virtues that provide necessary flexibility to rigidness of legal rules.[19]

The challenge goes deeper than doctrinal *de lege lata* and *de lege ferenda* research is able to answer, calling for a critical systemic reassessment of law's deep structure, building on interdisciplinary foundation. Taking into consideration the existing research on algorithmisation, are there alternative ways to conceptualise legal decision making and the role of human decision-makers? In turn, such analysis requires us to understand how the current human-faced procedures have come about and how these forms encompass different historical layers of regulatory design and the gradual social evolution over past decades of computers and centuries of fair trial principles. Without systemic overview, we face the risk of throwing the baby out with the bath water, of getting rid of vital components of procedural justice simply because we do not recognize them as such. Rouvroy argues that *prima facie* redundancies and inefficiencies of court procedure may in fact be constitutive for the use of discretionary power. By creating silences and pauses, breaks in the process pipeline, inefficiency may be necessary for creating a space for discretion, to support decision-makers' intuitive thinking.[20] In a similar vein, in addition to easy implementation into technological design, we need to critically examine what makes human oversight valuable, and what the procedural values and principles attached to human decision-makers that constitute due process and procedural justice are.

Hence, systemic review is needed but remains difficult, as it is not always clear what is needed. Against this backdrop, what would human oversight entail, if it were to become the focal procedural mechanism to ensure legal compliance and protection of fundamental rights? In legal decision making, such functions are allocated to the human decision-makers, which include judges in their courts, case workers and bureaucrats in public administration, parliamentary ombudsmen and other authorities in charge of legality control, specifics depending on the context and constitutional design. Legal institutions as we know them are often defined by their human actors, and legal decision making is defined by the discretionary power exercised by the personalised representative of the institution, the judge or administrator. The importance of discretion is reflected in the difficulties of automating legal decision making as well as differentiating such hard cases from routine cases that are allegedly more pliable to automation. Algorithmisation enables us to examine in new ways, whether and why this human element is necessary, and what would be lost if human decision-makers become human overseers.

In the following section, I examine how human oversight is conceptualized in the EU's emerging AI policy and what objectives are assigned to it, to elaborate on what meaningful human oversight should be in legal decision making.

---

19. K. Yeung, in K. Yeung and M. Lodge (eds.) *Algorithmic Regulation*, p. 29. The human judge's ability to exercise common sense has also been framed as the inductive element of legal decision making, where human input provides flexibility and renewal of law and deductive logic of legislation provides legal certainty. See e.g. A. Cornelis, 'Is it possible to program an ethical system? Foundation of Social Impact Assessment in the Theory of Informatics', in A. Martino, F. Natali and S. Binazzi (eds.), *Automated Analysis of Legal Texts. Logic, Informatics, Law* (Elsevier, 1985), p. 53.

20. A. Rouvroy, 'The end(s) of critique: data behaviourism versus due process', in M. Hildebrandt and K. de Vries (eds.), *Privacy, Due Process and the Computational Turn: The philosophy of law meets the philosophy of technology* (Routledge, 2013), p. 151.

Human oversight in EU regulation and policy

Many issues related to increasing ADM deployment are framed through the EU's General Data Protection Regulation (GDPR, 679/2016)[21] which, since its implementation in May 2018, has provided the primary European legal instrument for automated data processing, data protection and privacy. According to Article 22 on automated decision making and profiling, a data subject has the right not to be subject to a decision based solely on this automated processing, which produces legal effects concerning him or her or similarly significantly affects him or her. Although the article suggests a ban on automated decision making, the ban's importance is narrowed down by the broad exceptions.[22] For example, automated decision making is rendered compliant with GDPR if it is mandated by national legislation and enough measures are ensured to safeguard the data subject's rights and freedoms and legitimate interests Article 22(2b). Simply put, Article 22 of GDPR sets the stage for human oversight over algorithmic systems.

Before being replaced by the European Data Protection Board since the implementation of the GDPR, the independent advisory body called the Article 29 Working Party (WP29) produced guidelines on the interpretation of the regulation's Article 22.[23] The WP29 guidelines define automated decision making as 'the ability to make decisions by technological means without human involvement', which may or may not include or overlap with profiling, which refers to 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements'.[24] The WP29 guidelines emphasise that fabricated human involvement is not sufficient to avoid the Article 22 provisions, but instead, the human oversight needs to be meaningful in the sense that the overseer should have the authority and competence to change the decision.[25] Also, the procedural safeguards stipulated by the GDPR highlight the data subject's ability to contest the processing of their data, stimulating discussions on how to implement contestability to technological design.[26]

21. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), [2016] OJ L 119.
22. See also M. Brkan & G. Bonnet, 'Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas', 11 *European Journal of Risk Regulation* (2020), p. 18.
23. 'Guidelines on Automated individual decision making and Profiling for the purposes of Regulation 2106/679, wp251rev.01', *Article 29 Data Protection Working Party* (2018), https://ec.europa.eu/newsroom/article29/item-detail.cfm? item_id=612053.
24. Ibid. p. 6-8.
25. Ibid. p. 21. Similarly, in legal scholarship, Brkan addresses the performative human control in 'rubber stamping' and contends that meaningful intervention by human overseer requires both authority and capacity to change the decision based on automated processing. See M. Brkan, 'Do Algorithms Rule the World? Algorithmic Decision making and Data Protection in the Framework of the GDPR and Beyond', 27 *International Journal of Law and Information Technology* (2019), p. 91, 94.
26. See, for example, M. Almada, 'Human intervention in automated decision making: Toward the construction of contestable systems', *17th International Conference on Artificial Intelligence and Law (ICAIL 2019)* (Forthcoming), pp. 2-11.

The Commission's new digital package published in February 2020 elaborates the future policy setting around AI and other data-intensive technological systems.[27] The digital package includes four documents, all of which echo the EU's established agenda about 'trustworthy' and 'human-centric' AI, built on the acknowledged importance of human oversight for the protection of human autonomy and fundamental rights.[28] For mapping out the scope of human oversight, particularly interesting are the Report on the safety and liability implications and the White paper on AI.

The Report on the safety and liability implications of AI, Internet of Things and robotics locates the challenges of AI systems within the safety and product liability framework, which are regulated in the EU through the General Product Safety Directive.[29] Thus policy issues are framed in terms of consumer protection, overall trust and predictable regulatory environment for businesses, suggesting that AI-related problems can be solved through these regimes. Simultaneously, the focus also narrows down the problem representation. Although the EU agenda understandably focuses on the policy issues that can be solved within the scope of the EU's mandate, the problem definition excludes the cumulative and collective effects of algorithmisation that fundamentally challenge existing legal concepts and structures. Furthermore, limited focus excludes other legal fields and their mechanisms from potential policy action, although approaches from competition law, labour law, tax law and procedural law might bring additional instruments to the policy debate.

The Report describes the characteristics of AI that give rise to the policy issues. These include the connectivity, autonomy and data dependency that enable AI systems to 'perform tasks with little or no human control or supervision'.[30] The Report construes the so-called 'black-box effect' as a central policy problem, referring to the opacity of some AI systems,[31] that can be solved by transparency and explainability. Policy actions could include obligations on developers to disclose design parameters and metadata of datasets in case of accident and *ex post* assessment conducted by enforcement authorities based on such information disclosure. Finally, the Report implies that human oversight may become a focal procedural safeguard for controlling AI systems throughout their lifecycle, with specific requirements being implemented in EU legislation.[32] If implemented, this stance sets high expectations for the efficiency of human oversight.

The White paper on AI perceives AI harms both as material (for example health, damage to property) and as immaterial (for example privacy, human dignity discrimination),[33] recognising how AI deployment can aggravate such patterns 'without [the presence of] the social control

27. Report from the Commission to the European Parliament, the Council and the European economic and social committee, Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM (2020) 64 final; Commission White Paper Artificial Intelligence - A European approach to excellence and trust, COM (2020) 65 final; Communication from the Commission to the European parliament, the Council, the European economic and social committee and the Committee of the regions, A European strategy for data, COM (2020) 66 final; Communication from the Commission to the European parliament, the Council, the European economic and social committee and the Committee of the regions, Shaping Europe's digital future, COM (2020) 67 final.
28. "The Commission's vision stems from European values and fundamental rights and the conviction that the human being is and should remain at the centre." See COM (2020) 66 final, p. 4.
29. COM (2020) 64 final, p. 2.
30. COM (2020) 64 final, p. 2.
31. COM (2020) 64 final, p. 9.
32. COM (2020) 64 final, p. 8.
33. COM (2020) 65 final, p. 10.

mechanisms that govern human behaviour'.[34] To combat these risks, the white paper proposes the adoption of a risk-based approach, based on a two-pronged definition of high risk. An *ex ante* risk assessment should be performed before AI deployment in high-risk sectors, in which significant risks may occur such as in public administration, and for intended uses when risks may occur.[35] The White paper explains that the mandatory legal requirements for AI need to be decided as a part of the design of the regulatory framework.[36] In short, legal decision making should be considered to be a high-risk domain in which hard law regulation on AI may be necessary. Here, human oversight becomes a potentially important regulatory tool.

The key requirements for high-risk AI applications listed high in the White paper are the recommendations of the High-Level Expert Group on AI, which include human oversight.[37] The White paper also connects the definition of high-risk sectors and uses, which include the AI applications for a specific legal regime, and the required level of human oversight with implementation examples.[38] Firstly, human oversight may be required to review and validate the system's output before the decision becomes effective. Thus, the rejection of a social security application would always be taken by a human. Secondly, human oversight might take the form of human review after the decision has become effective, enabling human intervention when needed. The given example involves automated rejection of credit card application. Thirdly, human oversight could be implemented to monitor the AI system while in operation and the ability to intervene, by including a stop button or procedure. Fourthly, certain operational constraints can be imposed in AI design. The example for design-based constraints is found in driverless cars that switch control to humans in certain situations.

The EU's emerging AI policy seems to focus on *ex ante* assessment, at the expense of marginalising the elaboration of *ex post* mechanisms. This approach might clash with the *ex post* mechanisms, based on which legal protection typically is produced, although the topic deserves more thorough analysis. It remains unclear how the *ex ante* emphasis is able to provide solutions to acknowledged lack of efficient redress.[39] The policy documents propose human oversight in its different forms as the primary *ex post* mechanism to ensure legal protection. However, there is little substance by itself to the mechanism, leaving open what needs to happen between the AI system and the human overseer to constitute as meaningful protection. It is assumed that human oversight provides protection, partly following the GDPR's framing that grants it intrinsic value as a procedural mechanism that justifies automated processing.

34. COM (2020) 65 final, p. 11.
35. 'For instance, uses of AI applications that produce legal or similarly significant effects for the rights of an individual or a company; that pose risk of injury, death or significant material or immaterial damage; that produce effects that cannot reasonably be avoided by individuals or legal entities.' See COM (2020) 65 final, p. 17.
36. COM (2020) 65 final, p. 8.
37. Independent High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI', *European Commission* (2019), https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. See also Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee and the Committee of the Regions on Building Trust in Human-Centric Artificial Intelligence, COM (2019) 168 final.
38. COM (2020) 65 final, p. 21.
39. K. Yeung, in K. Yeung and M. Lodge (eds.), *Algorithmic Regulation*, p. 29.

Human oversight and the (false) promise of control

The importance of human oversight over automation is widely acknowledged in socio-legal scholarship. For example, John Danaher contends that because reliance on ADM limits active human participation, the systems impose a fundamental threat to legitimacy, which he considers difficult to accommodate or resist. In her work on law, technology and philosophy, Mireille Hildebrandt addresses similar issues of justification and discusses the need for protection of 'what is uncountable, incalculable or incomputable about individual persons', which comes under threat in the context of automated decision making, in which contestation by those subjected to automation plays a vital role.[40] It seems that human participation is required of decision-makers and of those subjected to the consequences of decisions. Of these forms of participation, human oversight relates primarily to the role of the human decision-maker, whose task it is to ensure also meaningful participation to those affected. Hence, this role would entail imposing the ultimate control over algorithmic systems, ensuring the protection of material and procedural rights.

Contrary to the promise associated with human oversight in policy making, the practical limitations of human capabilities to control complex technological systems have been much discussed. At times, the regulatory frameworks and technical standards contribute to assigning blame to the human-in-the-loop, practice conceptualised by Elish as 'a moral crumple zone to describe how responsibility for an action may be misattributed to a human actor who had limited control over the behaviour of an automated or autonomous system'.[41] For various reasons, from boredom at routine monitoring to automation bias and alert fatigue, humans generally perform badly as supervisors of automated technical systems.[42] Sociologist John Perrow discusses human error when accidents happen in complex technical systems and argues that the combined effects of tightly coupled complex systems and a high risk potential render accidents unavoidable by simple design choices.[43] Science and Technology Studies scholar Sheila Jasanoff discusses the fabricated and performative nature of 'human pretensions of control over technological systems'.[44]

These insights reveal how the feasibility of human oversight needs to be questioned in policy making. Still, despite its limitations, human oversight becomes an enticing option for procedural protection because it can be operationalized relatively easily, both within the technical specification and within the legal system. Human oversight follows the form of institutionalized decision making processes as they currently are, defined by human input, and in this sense its adoption as the definitive procedural safeguard would not signify a far-reaching change of existing conceptualisations. Furthermore, through digital traces and metadata we can easily ascertain that the oversight requirement has been fulfilled, that a human user has indeed been in the loop, on the

40. M. Hildebrandt, 'Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning', 20 *Theoretical Inquiries in Law* (2019), p. 83-121. See also, M. Hildebrandt and K. de Vries (eds.), *Privacy, Due Process and the Computational Turn: The Philosophy of Law meets the Philosophy of Technology.*
41. M. Elish, 'Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction', 5 *Engaging Science, Technology, and Society* (2019), p. 40.
42. For example, Bainbridge points out that 'by taking the easy part of his task, automation can make the difficult parts of the human operator's task more difficult'. See, L. Bainbridge, 'Ironies of Automation', 19 *Automatica* (1983), p. 775-779. See also, K. Yeung, in K. Yeung and M. Lodge (eds.), *Algorithmic Regulation*, p. 25.
43. J. Perrow, *Normal Accidents: Living with High-Risk Technologies* (Basic Books, 1984).
44. S. Jasanoff, 'Technologies of Humility: Citizen Participation in Governing Science', 41 *Minerva* (2003), p. 223.

loop or in command, to have pressed the button, or ticked the box. For these reasons, human oversight may easily become a value, despite its limited feasibility.[45]

Algorithmisation seems to pose a new form of architectural control that becomes entwined with legal structures, suggesting that it is the technological design that becomes the object of human oversight. As Cohen states, legal institutions are not fixed but instead influence and are influenced by managerial economic logic and affordances of technological structures.[46] Yeung and Lodge emphasise the cumulative effect of legal and technological normativity, where the technological design is applied for regulatory purposes, creating new forms of control.[47] This normativity of technological architecture has long been recognised, some declaring it *sui generis*[48] and others arguing for a more generic concept of normativity that would not differentiate between regulatory and technological design.[49] Still, the legal system continues to struggle to include new computational tools in its existing structures and operations, as can be exemplified by a recent study on the evidentiary value of ADM outputs in the courts.[50]

Decreasing human discretion in face of the computational turn

The challenges to law posed by algorithmic decision making do not boil down to issues of data governance but instead are related to even more fundamental societal shifts, the increasing importance and reliance on computational processes in organising social interaction or, in Hildebrandt's terminology, the computational turn.[51] By this framing, we are able to perceive algorithmisation in its historical context, which opens up new venues for examination. When algorithmisation is seen as one of the most recent steps of a continuum defined by the computational turn, we are able to overcome a typical thought error related to technological change, as we tend to overexaggerate the short-term and underestimate the long-term implications of any new technological tool. Such contextualisation not only fixes the lack of long-term perspectives often associated with new digital technologies but also opens up the lessons learned from early adoption of computers in legal and administrative processes since the 1970s. In this sense, algorithmisation takes place against the backdrop of long-term development of evidence-based decision making defined by quantification and data intensity, where governance is increasingly based on the establishment of measurable indicators and the application of statistical methods.[52] This contextualized approach helps us to locate how algorithmisation is influencing legal decision making in the long term and draw from earlier research to understand better what happens when human decision-makers are becoming human overseers of algorithmic systems.

45. K. Yeung, in K. Yeung and M. Lodge (eds.), *Algorithmic Regulation*, p 25; R. Koulu, *European Journal of Legal Studies* (2020) (forthcoming).
46. J. Cohen, *Between Truth and Power*, p. 5 and 143.
47. K. Yeung and M. Lodge (eds.), *Algorithmic Regulation*, p. 5.
48. See e.g. L. Lessig, *Code: Version 2.0* (Basic Books, 2006), http://codev2.cc/download+remix/.
49. M. Hildebrandt, 'Legal and Technological Normativity: more (and less) than twin sisters', 12 *Techné* (2008), p. 169.
50. G. Vanderstichele, 'The Normative Value of Legal Analytics. Is There a Case for Statistical Precedent?', *Master Thesis University of Oxford* (2019), https://ssrn.com/abstract=3474878.
51. M. Hildebrandt and K. de Vries (eds.), *Privacy, Due Process and the Computational Turn: The Philosophy of Law meets the Philosophy of Technology.*
52. See, for example, K. Davis, B. Kingsbury and S. Merry, 'Introduction: Global Governance by Indicators', in Kevin Davis et al. (eds.), *Governance by Indicators. Global Power through Quantification and Rankings* (Oxford University Press, 2012), p. 3-28.

Early research on computerisation of public administration also demonstrates the importance of the organisational setting in which technological systems are deployed. These perspectives highlight how the long-term consequences are context-dependent, which also affects the role reserved for human overseers. The unique implementation characteristics define the results of deployment, such as the adoption rate and user resistance. Such characteristics are highly context-dependent yet define the overall objectives and prioritised functionalities of the developed system: the hegemonic actors in the organisation and the deployment process, the formal and informal practices established in the organisation, the composition of the development team and the maintenance control.[53]

In empirical work on computers in public administration, Kenneth Laudon contextualised their adoption through managerial rationality that emerged in the US from the Progressive reform in the 19th century that built on the ideal of rational effective government.[54] Laudon also elaborated the political dimension of computerised bureaucratic reform: to administrative reformers of the 1960s and 1970s, computers signified a step closer to the bureaucratic ideal of rational effectiveness, without having to risk a gruesome political debate on political objectives or a redistribution of political power.[55]

Additionally Laudon emphasised centralisation in computerisation, the loss of control from the public organisation to the central data bank, placing centralisation as one central variable related to increasing reliance on automated information processing.[56] Laudon's study also demonstrated that if the objectives are defined by a central authority external to the implementation organization, the deployment is likely to focus on cost savings, increasing central control and employee surveillance, whereas objectives defined internally by the organisation itself tended to focus on improving the quality of internal processes.[57] Centralisation as well as standardisation of work practices are the necessary prerequisites to establish a standardised process outline which can then be automated.

Similarly, Shoshana Zuboff describes automation in terms of a shift from comprehensive labour requiring artisanal expertise to fragmented tasks requiring no prior skills, documenting how this dumbing down has a negative impact on employees' sense of control over their work and performance.[58] Lisa Bainbridge notes the 'ironies of automation', when the technology implementation does not necessarily benefit the human operator but instead makes their work more difficult by removing the routine tasks and allocating the harder monitoring and take-over tasks to them.[59] For our discussion, these insights document how human labour and sense of control changes when humans become overseers of automation, suggesting that similar loss of control might also be experienced by legal decision-makers faced with increasing algorithmisation.

Often overlooked particularly in doctrinal legal studies, the complexity and context-dependency that result from institutional design are necessary for understanding the long-term consequences of

53. For example, Laudon describes how the organisation's internal and external integration affect the consequences of computer deployment. In organisations with high internal integration between employees, there was a stronger likelihood of resistance against computerisation. See K. Laudon, *Computers and Bureaucratic Reform. The Political Functions of Urban Information Systems* (John Wiley & Sons, 1974), p. 73.
54. Ibid. p. 41-61.
55. Ibid. p. 52.
56. Ibid. p. 65.
57. Ibid. p. 308.
58. S. Zuboff, *In the Age of the Smart Machine. The Future of Work and Power* (Basic Books, 1988), p. 48.
59. L. Bainbridge, 19 *Automatica* (1983), p. 775-779.

algorithmisation for legal decision making. At least three observations can be made here. Firstly, algorithmic decision making constitutes a type of 'non-decision making' in the sense that existing institutional processes are standardised, datafied and automated by establishing the decision making parameters in advance and implementing them in the architectural system design. Simply put, the decision making is moved upstream, to the system architects and the boundary work between the domain and technology experts, further away from the 'street-level bureaucrats'.[60] Despite the context-dependency of technology deployment, loss of human discretion at the grass-roots level has been described as a central consequence of the computational turn. Secondly, this loss of discretion should be examined in relation to the discretionary space reserved to human decision-makers by the institutional and regulatory design. We should not assume that human decision-makers were free to use their discretionary freedom before the computational turn, as particularly in public administration, decision-makers are required to follow not only the legislation but also lower decree regulation as well as internal guidelines and organisational practices which traditionally have limited the scope of discretion significantly. Thirdly, it is likely that existing organisational practices of legal institutions are defined by human tasks, completed at most with the help of simple office automation tools. These human-oriented tasks also form the model for technology deployment by establishing the process, which is then implemented in the advanced technological systems without systemic overview of their functions.

These observations highlight the importance of examining the human tasks that constitute legal processes, as these are implemented both in the normative materials such as legislation and to the informal everyday practices of administrating justice. Algorithmisation calls for empirical research to map out the everyday decision making practices that become increasingly performed with technological systems. In short, we cannot conceptualise algorithmisation only by looking at law in books but instead the focus needs to be on algorithmisation of law as it is in action. For such analysis, the socio-materiality of technological forms provides a complementary perspective.[61]

By pointing towards the everyday encounters human decision-makers have with technological systems, socio-materiality reveals the ways in which decision-makers use digital technologies in their everyday work and how their perceptions shape the consequences of algorithmisation.[62] The research conducted on computer-supported collaborative work and human-computer interaction might also provide new venues for socio-legal analysis, enabling us to ask about the extent to which extent improved interaction design can mitigate the loss of discretion that follows from standardization and automation.

60. See, for example, A. Alkhatib and M. Bernstein, 'Street-Level Algorithms: A Theory at the Gaps Between Policy and Decisions', *CHI* (2019); J. Pääkkönen et al., 'Bureaucracy as a Lens for Analyzing and Designing Algorithmic Systems', *CHI* (2020).
61. See, for example, W. Orlikowski, 'Sociomaterial Practices: Exploring Technology at Work', 28 *Organization Studies* (2007), p. 1435–144.
62. If we are to examine how the sociomateriality of technology influences legal decision making processes, it is also necessary to acquire empirical knowledge about the everyday work practices of lawyers, public administrators and legal decision makers becoming increasingly defined by digital technologies. Some such studies exist. For example, the Law Society of England and Wales has collected information about lawyers' technology use through online surveys and interviews. Although the results cannot be generalized across jurisdictional and cultural contexts, they give an overview of on-going developments. See, 'Capturing Technological innovation in Legal Services', *Law Society of England and Wales* (2017), https://www.lawsociety.org.uk/support-services/research-trends/capturing-technological-innovation-report/.

To those seeking justice in the courts, procedural rights such as the right to be heard are defined through the encounter with the human judge. From the judge's perspective, institutionalised into procedural doctrine and due process standards, reason giving is a primary instrument to prove such an encounter has taken place. The written grounds of the judgment communicate to the parties that they have indeed been heard, that their opinions have been noted and thoroughly considered, even though (and especially when) the outcome goes against them. But from the parties' perspective, the experience of being heard does not necessarily require written grounds but the ability to speak one's mind, as the seminal work of Lind and Tyler on social psychology of procedural justice demonstrates.[63] Yet encounter is not always important and to a certain extent it can be simulated (and already is). In debates on automation of legal decision making we often tend to forget the human is already out of the loop, that courts and public agencies apply standardised solutions and produce routine-like decisions that do not require discretion.[64] The encounter and the discretion become decisive when creativity is needed, when there are exceptional circumstances at hand that need to be taken into consideration, and when the standard routine does not apply.

To conclude, the algorithmisation of legal decision making is a multifaceted, dynamic process heading towards hybridisation in complex socio-technical systems, which connects with shifting modalities of governmentality and the changing roles legal institutions and administrative organisations play in the society. In this sense, algorithmisation is not simply about technology deployment or even the limits of automation but about complex interdependencies between actors, regulatory frameworks as well as inter- and intra-organisational formal and informal practices that define how humans make legal decisions increasingly with the support of technological systems. The normativity of technological architectures becomes entwined with the long-term trends towards increasing standardisation, datafication and automation of decision making processes, contributing to the progressive loss of human discretion, raising concerns for the protection of due process and other fundamental rights we associate with human discretion. In short, human discretion is diminishing — or moving further away from the grassroots level — with technological architectures, perhaps increasingly so with more autonomous algorithmic systems. It is possible that assigning human decision-makers the role as human overseers is contributing to this development instead of mitigating its negative consequences, if human oversight becomes performative rubberstamping without meaning.

For the critical observer, algorithmisation reveals how law construes fairness through the human element of legal decision making, how due process standards are formulated around tasks and capabilities of human actors. Broad discretion tempered by accountability structures introduces freedom to decision making, but this freedom is not random or arbitrary but instead creates space for contextual consideration, understanding for the exceptional circumstances unique to that case. Similarly, the experience of being heard is built on the impression of human encounter between those seeking justice and the face of justice embodied in the person of the human decision-maker. Discretion and the encounter can be monitored through the human actor's reason

63. T. Lind and A. Tyler, *The Social Psychology of Procedural Justice* (Plenum Press, 1988).

64. In the context of court IT and Dutch civil procedure, Dory Reiling argues that the development of court technology often builds on assumptions instead of the reality of the court organisation. Contrary to what the focus on physical court proceedings, the day in court, might suggest, majority of civil cases are resolved in written procedures that require very different digital tools than full-scale trials. Hence systemic review of what the courts actually do is necessary for successful technology implementation. See D. Reiling, *Technology for Justice. How Information Technology Can Support Judicial Reform* (Leiden University Press, 2009), p. 116, http://home.hccnet.nl/a.d.reiling/.

giving that ultimately justifies what has been decided and why. To provide justification for exceptional circumstances, reasons cannot be general clauses but instead need to reflect that what has been heard has also been considered.

Conclusion

It seems more or less surprising that algorithmisation makes visible the necessary human element of legal decision making, which is always implicitly present in legal materials but seldom explained in legal thought that tends to focus on legal certainty and objectively measurable equality before the law. In this sense, algorithmisation provides us with a steppingstone to map out the underlying genealogy of modern legal processes and critically assess the basic assumptions, values and objectives that are embedded in procedural design. For example, algorithmisation reveals the personalized dimension of legal decision making, reflected in articulations of human oversight advocated in EU policy and the right to a human judge promoted by CEPEJ's ethical guidelines on AI use in the judiciary.

The value given to human judgement demonstrates that we attribute to humans the ability to produce legitimacy and justification for decisions, a task we perceive fundamentally impossible for machines. We notice it is not enough to strive for legal certainty and objective, predetermined rules but instead fair decision making requires creativity and freedom produced by the human individuals. However, this also creates unpredictability.

Against this background, the personalisation of justice hidden in law's deep structure becomes visible: legal decision making is framed through the actions of human decision-makers, most importantly through discretion. On one hand, discretion is arbitrariness that needs to be tamed in through accountability mechanisms, institutionalised in principles on the judge's strict personal responsibility, publicity and transparency, sanctions for misuse of power. On the other, it is creative freedom that needs to be enabled, institutionalised in principles on the independence and impartiality of the judge. This tension between freedom and control, definitive of legal decision making and sedimented in legal thought as the personal dimension of the human decision-maker, is revealed by the mirror of algorithmisation. The tasks traditionally performed by humans (judges, clerks, case workers and bureaucrats) are embedded in conceptualisations of due process and procedural justice. This observation suggests that automation should not be focused on automating human tasks, but instead the objectives and functions they reflect, which in turn requires an in-depth understanding of legal decision making.

Another reflection surface is provided by the complex body of literature on the organisational embeddedness and socio-materiality of technology, and hybridisation in complex socio-technical systems that gives context to the algorithmisation of legal decision making. It demonstrates that the consequences depend on various factors (of which formal legal structures are but one) making it difficult to obtain a comprehensive overview of long-term implications. However, when conceptualised in terms of the on-going computational turn or digitalisation of legal practices, the worrying loss of discretion comes into focus. It becomes clear that more research, both theoretical and empirical, is needed to understand better the how the tension between the personal dimension and organizational context of legal decision making plays out with algorithmisation.

This analysis provides a contribution to a more comprehensive view of the procedural dimension of algorithmisation. The examination locates the risks associated with AI deployment in the legal domain to the focal role granted in law to the human decision-maker's discretionary power. Discretion as freedom and control is the definitive logic of the way legal processes operate and how

law is renewed through precedents based on errors, conflicts, exceptions and the creative use of discretionary power. If we are to impose control over AI systems through human oversight, we need to ask what makes it meaningful and for that purpose we need an understanding of legal decision making, both as a normative ideal and as collective everyday practices. The importance of the human element for law sets high expectations for human oversight. It is unlikely that despite the best intentions, that human oversight as such would provide a solution to the negative consequences of algorithmisation. Here the EU's emerging AI policy provides little guidance for what constitutes 'meaningful'. Furthermore, the proven limitations of humans as overseers of automated systems cautions against attaching too much expectation to its feasibility. If we are not cautious of these limitations, human oversight faces the risk of becoming a value in itself, turning human decision-makers into rubberstamps, providing a sound procedural mechanism that can easily be verified after it has been conducted yet being empty of meaning. In this sense it is possible that emphasis on human oversight may aggravate the long-term negative consequences the computational turn is having on discretion.

Finally, we return to the double meaning already embedded in the concept of human oversight. Oversight can be failure or mistake, so definitive of human action, as well as the action of overseeing, ensuring that failures or mistakes do not happen. Thus, equating the problem and the solution, the concept encapsulates the paradoxical human dimension of decision making, and the continuous balancing between freedom and control, discretionary power and legal certainty.

ORCID iD

Riikka Koulu https://orcid.org/0000-0002-1298-2406